

General Goals Update NIST guidelines to: (i) *increase freedom of choice* required by scientists, statisticians, and mathematicians to be able to address needs of rapidly evolving and expanding fields of measurement science; (ii) *Widen class of measurement models* used to assign values to measurands and to evaluate measurement uncertainty; (iii) *Facilitate critical assessment of models and assumptions*, in particular of those that support the probabilistic interpretation of measurement uncertainty.

Grandfathering All uncertainty evaluations published as part of measurement results produced in the delivery of NIST measurement services (reference materials and calibrations) remain valid and need not to be redone; the uncertainty evaluation procedures described in NIST TN 1297 and in the GUM (1995, 2008) may continue to be used going forward.

Measurement is understood in a wide sense as an experimental or computational process that produces a measurement result intended for use in support of decision-making (*cf.* R. White, 2011, ACQUAL 16: 31–44). A measurement result comprises: (i) an estimate of the value of a property of a material or virtual object or collection of objects, or of an event or series of events, using empirical data possibly in conjunction with preexisting knowledge; and (ii) an evaluation of the measurement uncertainty associated with that estimate.

Measurement uncertainty is a quantity that characterizes the dispersion of the values that may be attributed to a measurand and that are consistent with the empirical data and with other relevant information about the true value of the measurand.

- A probability distribution, on the set of values of the measurand, that describes the state of knowledge about the true value of the measurand, provides the most complete characterization of measurement uncertainty. This state of knowledge may be an individual scientist's, or a scientific community's.
- In many cases, and for scalar measurands, the dispersion of values may be summarized by the standard deviation of this distribution (*standard uncertainty*). For multivariate and more general measurands, analogous summaries may be used.
- For nominal properties, the entropy of the corresponding distribution is a summary description of measurement uncertainty.

Measurement models describe the relationship between the value of the measurand and the quantities used to estimate it: (i) a *measurement equation* expresses the measurand as a known function of a set of input quantities; (ii) an *observation equation* (or, *statistical model*) expresses the measurand as a known function of the parameters of the probability distribution that describes the variability of the empirical data used in measurement.

Uncertainty evaluations of **Type A** involve the application of statistical methods to experimental data, consistently with the measurement model. Evaluations of **Type B** involve the elicitation of (selected attributes of) probability distributions that describe states of knowledge about the values of participating quantities, for example using the *MATCH Uncertainty Elicitation Tool* available at optics.eee.nottingham.ac.uk/match/uncertainty.php.

Uncertainty propagation for measurands defined by measurement equations, use either Gauss's formula (Equation (A-3) in NIST TN 1297) and all the methods described in the GUM (1995, 2008) and in NIST TN 1297, or Monte Carlo methods described in the GUM Supplement 1: both are implemented in the *NIST Uncertainty Machine* (stat.nist.gov/uncertainty), with user's manual at www.nist.gov/itl/sed/gsg/uncertainty.cfm. For

measurands defined via observation equations, use methods of mathematical statistics, selected and applied in collaboration with a statistician or applied mathematician.

Express measurement uncertainty most completely by fully specifying a probability distribution that exactly or approximately describes the state of knowledge about the value of the measurand (both the *NIST Uncertainty Machine* and methods of mathematical statistics can produce arbitrarily large samples from such distribution). In many cases, a summary of the dispersion of values of the distribution suffices: for example, the standard deviation for scalar measurands, the covariance matrix for multivariate measurands (or the corresponding 68 % probability coverage hyper-ellipsoid), or more generally a *coverage region* that, with specified probability (95 % typically), is believed to include the true value of a scalar or multivariate measurand.

Example 1: Thermal Expansion Coefficient $\alpha = (L_1 - L_0)/(L_0(T_1 - T_0))$ of a copper bar whose lengths $L_0 = 1.4999$ m and $L_1 = 1.5021$ m were determined at temperatures $T_0 = 288.15$ K and $T_1 = 373.10$ K, with standard uncertainties $u(L_0) = 0.0001$ m, $u(L_1) = 0.0002$ m, $u(T_0) = 0.02$ K, and $u(T_1) = 0.05$ K. The estimate of the measurand is $\hat{\alpha} = 1.727 \times 10^{-5}$ K⁻¹, and Gauss's formula produces $u(\hat{\alpha}) = 1.8 \times 10^{-6}$ K⁻¹, both computed using the *NIST Uncertainty Machine*.

Example 2: Falling Ball Viscometer to measure the dynamic viscosity μ_M of a solution of sodium hydroxide in water at 20 °C, using a boron silica glass ball of mass density ρ_B . The measurement equation is $\mu_M = \mu_C[(\rho_B - \rho_M)/(\rho_B - \rho_C)](t_M/t_C)$, where $\mu_C = 4.63$ mPa s, $\rho_C = 810$ kg/meter³, and $t_C = 36.6$ s denote the viscosity, mass density, and ball travel time for the calibration liquid, and $\rho_M = 1180$ kg/m³ and $t_M = 61$ s denote the mass density and ball travel time for the sodium hydroxide solution. If the input quantities are modeled as independent Gaussian random variables with means equal to their assigned values, and standard deviations equal to their standard uncertainties $u(\mu_C) = 0.01\mu_C$, $u(\rho_B) = u(\rho_C) = u(\rho_M) = 0.5$ kg/m³, $u(t_C) = 0.15t_C$, and $u(t_M) = 0.10t_M$, then the Monte Carlo method of the GUM Supplement 1 as implemented in the *NIST Uncertainty Machine* produces: $\hat{\mu}_M = 5.82$ mPa s, $u(\hat{\mu}_M) = 1.11$ mPa s, and (4.05 mPa s, 8.39 mPa s) as approximate 95 % coverage interval for μ_M (Note that this interval is asymmetric relative to the estimate $\hat{\mu}_M$.)

Example 3: Characteristic Strength of alumina is measured using the observed stresses at which 32 specimens of the material fractured in a flexure test (J. B. Quinn and G. D. Quinn, 2010, *Dental Materials* 26: 135–147): 265, 272, 283, 309, 311, 320, 323, 324, 326, 334, 337, 351, 361, 366, 375, 380, 384, 389, 390, 390, 391, 392, 396, 396, 396, 396, 398, 403, 404, 429, 430, 435 MPa. A tenable statistical model (observation equation) describes the data as outcomes of independent random variables with the same Weibull distribution whose scale parameter is the characteristic strength. The maximum-likelihood estimate is 383 MPa with approximate standard uncertainty 7 MPa, and (369 MPa, 398 MPa) is an approximate 95 % coverage interval, computed using R package *bbmle*.

Example 4: Forensic classification of the source of glass fragments, based on a function built using mixture discriminant analysis (T. Hastie, R. Tibshirani, and J. Friedman, 2009, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, Springer), whose inputs are the mass fractions of oxides of the major elements (Na, Mg, Al, Si, K, Ca, Ba, Fe) and the refractive index, produces a probability distribution over the set of possible sources.

Given values of these mass fractions (%) 13.92, 3.52, 1.25, 72.88, 0.37, 7.94, 0, 0.14, and of the refractive index 1.51613, for a particular fragment whose source is unknown, the classifier produces the following output (discrete) probability distribution: building window (float) 0.38, building window (non-float) 0.55, vehicle window 0.07, containers 0, tableware 0, headlamps 0. The resulting assigned value is “building window (non-float)” because it is the most likely. The distribution may be used directly in subsequent Monte Carlo uncertainty propagations, or its dispersion may be summarized by its entropy $H = -(0.38 \log(0.38) + 0.55 \log(0.55) + 0.07 \log(0.07)) = 0.88$. Computations done using R package *mda* with data from K. Bache & M. Lichman (2013 *UCI Machine Learning Repository*, University of California, Irvine, CA).